Digital Footprinting, Analysis and Archival of Digital Media Content using Vision Transformers and Related Techniques

Sanket Anand
Student ID: 1140213
Thesis Supervisor: Dr. Kashav Ajmera

Final Thesis Presentation for
Master of Science in Computer Science

Liverpool John Moores University & upGrad

March 2024

Agenda

- Introduction
- Research Questions & Objectives
- Literature Review
- Research Methodology
- Implementation
- Results
- Conclusion and Future Works

Introduction

Artificial Intelligence (AI) is a broad concept that includes machine learning (ML) as a subfield. AI is the ability of computers to perform tasks like humans, while ML is the process of teaching machine to perform tasks by analyzing data.

Point Cloud is a discrete set of data points in space and is a common method of representing media data while preserving important properties such as permutation invariance.

Despite all the progress in recent years, surveys show that more than 80% of AI projects fail to achieve business goals due to multiple factors such as miscommunication with business, vendor dependence, etc.

These numbers highlight a major gap in the current ecosystem- lack of a standard framework or a design guideline that can be utilized to accurately bring together all the stakeholders and manage the complete lifecycle of an AI project.

This work proposes a framework for development and formal evaluation of an AI Implementation project that covers all the lifecycles of a generic AI project.

The framework will be evaluated in a generic context in order to evaluate its applicability to varying forms of data both qualitatively and quantitatively.

Research Questions & Objectives

Research Questions:

- Can we generalize the concept of a Machine Learning Model or a Machine Learning Project into a mathematical system?
- Can we accurately define an abstract representation capable of preserving and uniquely defining all the properties of any type of media content that we intend to feed into such a system?
- Can we generalize this system in the form of a framework that can be used as a guideline to build AI enabled software systems?
- How can we be absolutely sure that this framework is indeed strong enough to cover the class of all types of software systems possible?

Aim:

- This work attempts to generalize a machine learning system and analyze it from a mathematical lens whereby we discuss some generic guidelines for building an AI enabled software system.

Objectives:

- To conduct a comprehensive review of major advancements in Machine Learning and AI.
- To generalize the concept of a machine learning model into a mathematical system consisting of an abstract representation.
- To evaluate if this abstract representation is strong enough to represent the class of all software programs.
- To verify the applicability of this abstract representation into a generic AI enabled software system.

Challenges

Challenges in AI Development

- Miscommunication with business.
- Vendor dependence.
- Cost overrun.
- Performance issues.
- Inconsistent behavior such as bias, feature gaps, etc.

Challenges in Software Development

- High maintenance cost.
- Lack of flexibility.
- Staff upskilling is required to administer such a system.
- Longer downtimes.
- Can not work directly on unstructured data.

Literature Review – Abstract Structures

Group	Ring	Field	Vector Space
• The simplest of all abstract structures. • Used to denote a set with just one operation possible. • Some common examples include set of integers with addition, symmetry group, etc.	• Used to denote a set with two operations possible- addition and multiplication with addition both commutative and distributive, while multiplication just associative. • Example include ring of even integers.	• Field is a special type of ring that is commutative with identity and has multiplicative inverse for every element in it. • Some common examples are rational numbers, real numbers and complex numbers.	• A vector space over a field is a set of vectors and a field of scalars, along with two operators: vector addition and scalar multiplication. • For example, real vector space, complex vector space, etc.

Literature Review – Recent Progress made in AI

- Speech Processing
- Computer Vision
- Natural Language Processing

- Transformer Based Models
- Diffusion Models
- Generative Adversarial Networks (GANs)

- Code Copilots
- Task Operators
- Research Assistants

- Robotics
- Self Driving Cars
- Unmanned Space Vehicles

Literature Review – Point Cloud Processing

PointNet

- PointNet provides a unified architecture for applications ranging from object classification, part segmentation, to scene semantic parsing.
- Though simple, PointNet is highly efficient and effective.

PointNet++

- PointNet++ is a follow-up project that extends PointNet.
- It is the version 2.0 of the PointNet architecture.
- PointNet++ shows improvements from PointNet in the sense that it respects spatial localities and handles non-uniform densities well by a number of architectural improvements such as ability to learn hierarchical features with increasing scales of context like a CNN and special layers to aggregate information from different scales.

Dynamic Graph CNN (DGCNN)

- Dynamic Graph CNN achieves state-of-the-art performance on point-cloud related high-level tasks including category classification, semantic segmentation and part segmentation.
- It proposes a new neural network module dubbed EdgeConv suitable for CNN-based high-level tasks on point clouds including classification and segmentation.
- EdgeConv is differentiable and can be plugged into existing architectures.

Vector Neuron DGCNN

- Vector Neuron is a method of lifting standard neurons from R into the 3-dimensional space R*3.
- It easily integrates with the DGCNN as a backbone.
- The vector neuron model is SO(3) equivariant and is permutation invariant.

Literature Review – Theory of Computation

Finite State Machine

- A finite state machine (FSM) is a mathematical model that describes how a system can exist in a limited number of states.
- It is used to design and analyze systems with discrete and sequential behavior.
- It is an abstract machine that can be in exactly one of a finite number of states at any given time.

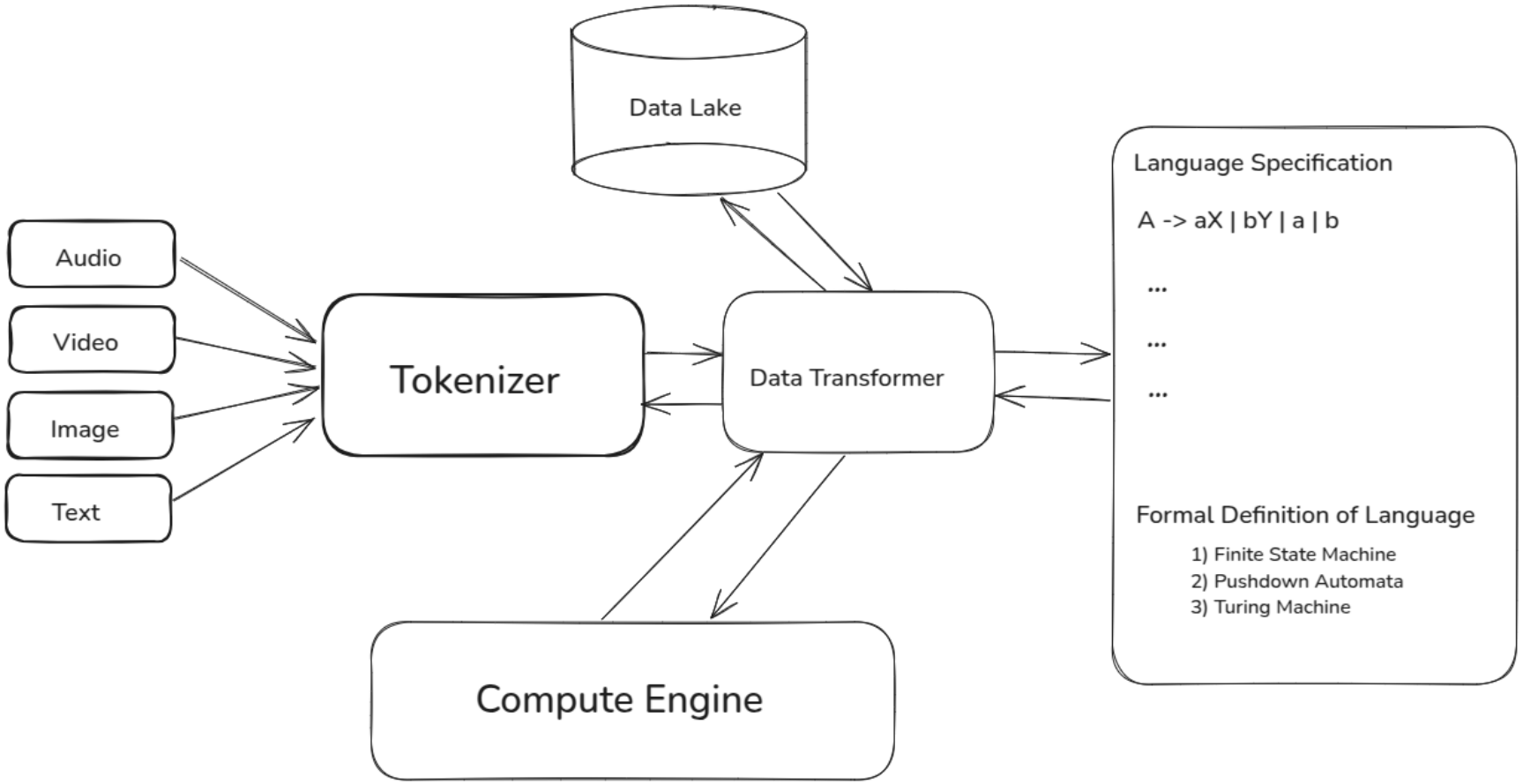
Pushdown Automaton

- A pushdown automaton, or PDA, is finite automaton supplemented with a storage device, specifically, a stack onto which we can push symbols.
- The transitions in the pushdown automaton read input symbols, just as a finite automaton does.

Turing Machine

- A Turing machine is a mathematical model of computation describing an abstract machine that manipulates symbols on a strip of tape according to a table of rules.
- Despite the model's simplicity, it is capable of implementing any computer algorithm.

Methodology – End to End Pipeline



Methodology - Algorithms & Techniques

Dynamic Graph CNN (DGCNN)- Dynamic Graph Convolutional Neural Networks (Dynamic Graph CNNs) are a type of neural network architecture designed to handle dynamic graphs. They extend traditional Convolutional Neural Networks (CNNs) by incorporating graph convolution operations that can process the changing structure of dynamic graphs.

Vector Neuron Network- Extending neurons from 1D scalars to 3D vectors, vector neurons enable a simple mapping of SO(3) actions to latent spaces thereby providing a framework for building equivariance in common neural operations – including linear layers, non-linearities, pooling, and normalizations.

Vector Neuron based DGCNN (VN-DGCNN)- It is an application of Vector Neuron Network applied to a Dynamic Graph CNN (DGCNN) backbone for classification operations.

Tree of Clarifications (ToC)- It recursively constructs a tree of disambiguation for the Ambiguous Query (AQ) via few-shot prompting leveraging external knowledge and uses it to generate long-form answer.

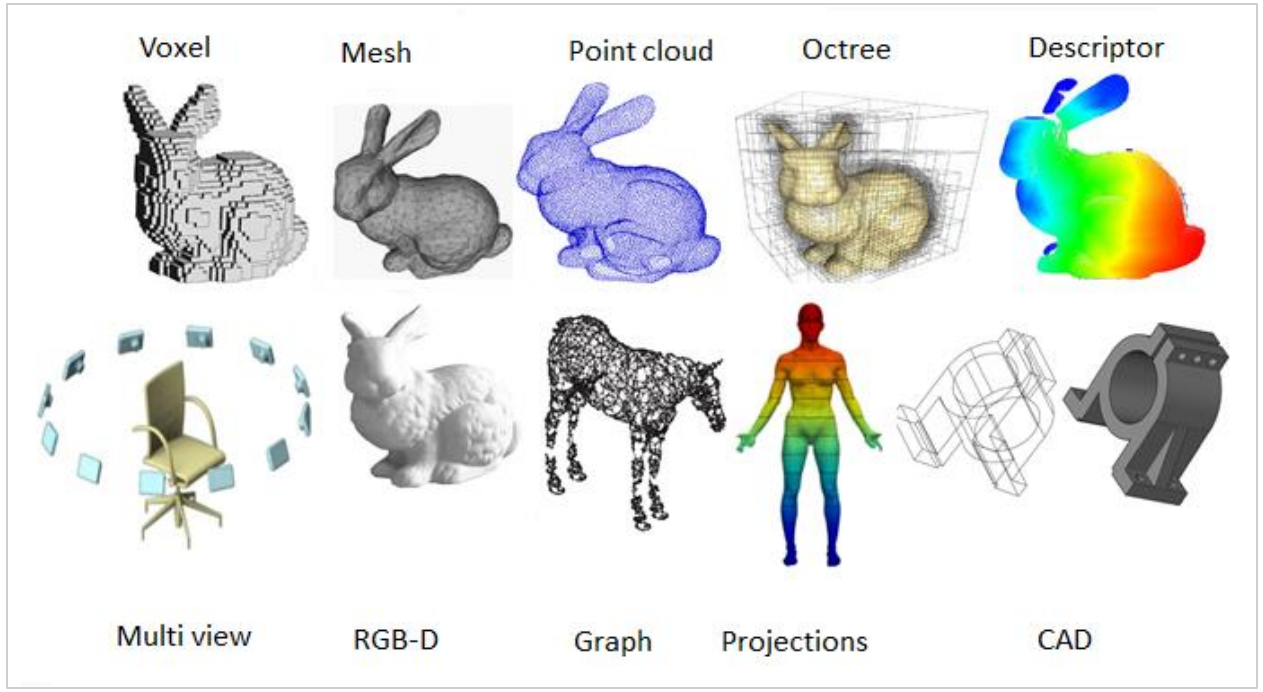
Church Turing Thesis- Every computation that can be carried out in the real world can be effectively performed by a Turing Machine.

Implementation – Data Description

ModelNet40 dataset- This dataset contains synthetic object point clouds. It is the most widely used benchmark for point cloud analysis. The dataset is specifically popular because of its various categories, clean shapes, well-constructed dataset, etc.

ASQA dataset- It is the first long-form question answering dataset that focuses on ambiguous factoid questions. Different from previous long-form answers datasets, each question is annotated with both long-form answers and extractive question-answer pairs, which should be answerable by the generated passage.

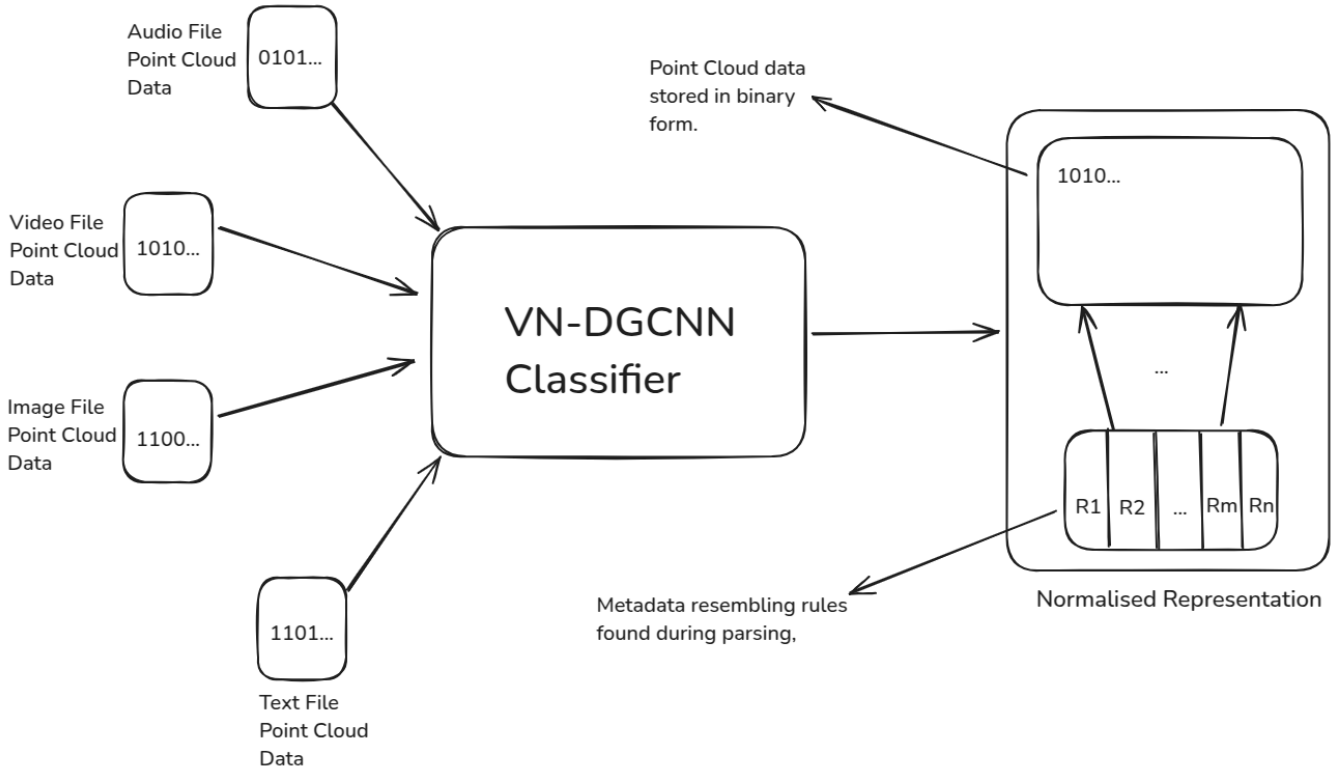
Implementation – Tokenizer



Point cloud and other variants (illustration)
(Gezawa A.S., et al., 2020)

- The image frames are tokenized in the form of point clouds using the python library “Open3D” to convert the RGBD (RGB with Depth Information) data from image frames into point clouds.
- Point Cloud is not the only option, choice of representation depends upon feasibility.

Implementation – Data Transformer

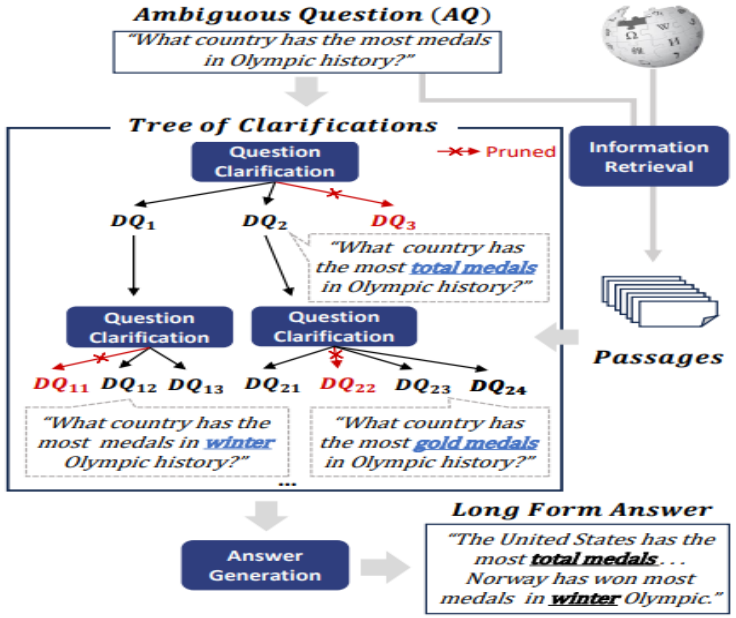


Data transformer implementation

- The transformer is responsible for validating and parsing the point cloud stream into a standard representation that includes metadata related to the classification done by VN-DGCNN.
- It essentially "lifts" the point cloud data into a common modality that can be analysed and worked upon.

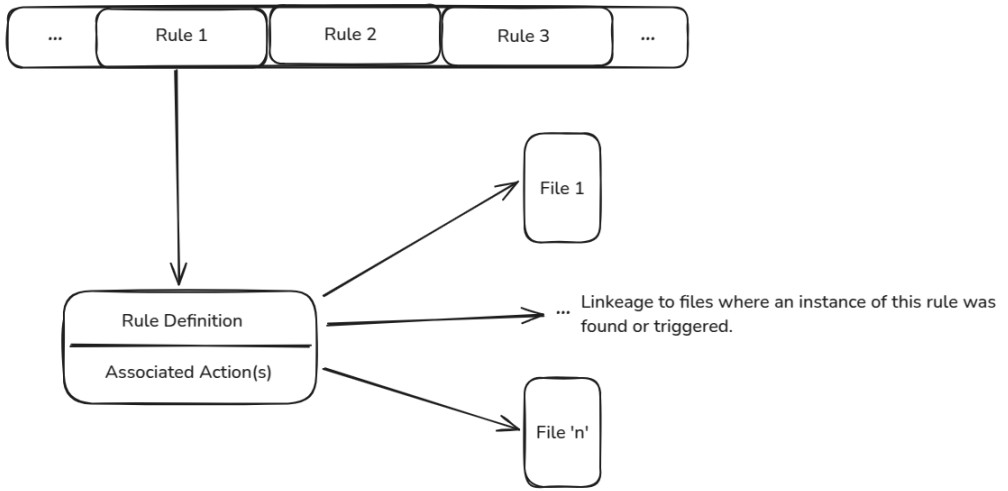
Implementation – Language Specification

Implementation – Compute Engine


Ambiguous Question (AQ)
"What country has the most medals in Olympic history?"
Tree of Clarifications
Pruned
Question Clarification
Information Retrieval
DQ_1
DQ_2
DQ_3
"What country has the most total medals in Olympic history?"
Question Clarification
Question Clarification
Passages
DQ_{11} DQ_{12} DQ_{13} DQ_{21} DQ_{22} DQ_{23} DQ_{24}
"What country has the most medals in winter Olympic history?"
"What country has the most gold medals in Olympic history?"
...
Long Form Answer
Answer Generation
"The United States has the most total medals... Norway has won most medals in winter Olympic."

Tree of Clarifications (Kim et. al., 2023)

Tree of Clarifications (ToC) framework is responsible for resolving user queries into disambiguated sub-queries and then combine the responses to return the answer to the ambiguated user query.



File directory structure

Tree of Clarifications framework essentially works bottom up by looking up all the grammar rules in the language specification relevant to the disambiguated sub-query. It consolidates the captured insights in that metadata to perform actions or return relevant response.

Results – Quantitative

Model Name	z/z	z/SO(3)	SO(3)/SO(3)
PointNet	85.9	19.6	74.7
DGCNN	90.3	33.8	88.6
VN-PointNet	77.5	77.5	77.2
VN-DGCNN	89.5	89.5	90.2

Classification accuracy on ModelNet40 dataset (VN-architecture) (Deng et. al., 2021)

- From (Deng et. Al., 2021), we can conclude that VN-DGCNN is SO(3) equivariant representation of point cloud data.
- It has been tested on classification operations to be permutation-invariant and rotation-invariant.

Results – Quantitative (Contd.)

Baseline	Model	Disambig-F1	ROUGE-L	DR Score
Fully Supervised	PaLM w/ Soft Prompt Tuning	27.8	37.4	32.1
Few-Shot Prompting (5-shots)	PaLM	25.3	34.5	29.6
Tree of Clarifications (ToC)	GPT-3 + RAC	31.1	39.6	35.1
Tree of Clarifications (ToC)	GPT-3 + RAC + TS	32.4	40.0	36.0
Tree of Clarifications (ToC)	GPT-3 + RAC + TS w/ Pruning	33.7	39.7	36.6

Evaluation comparison of long-form QA on ambiguous questions on ASQA dataset (Kim et. al., 2023)

Results – Qualitative

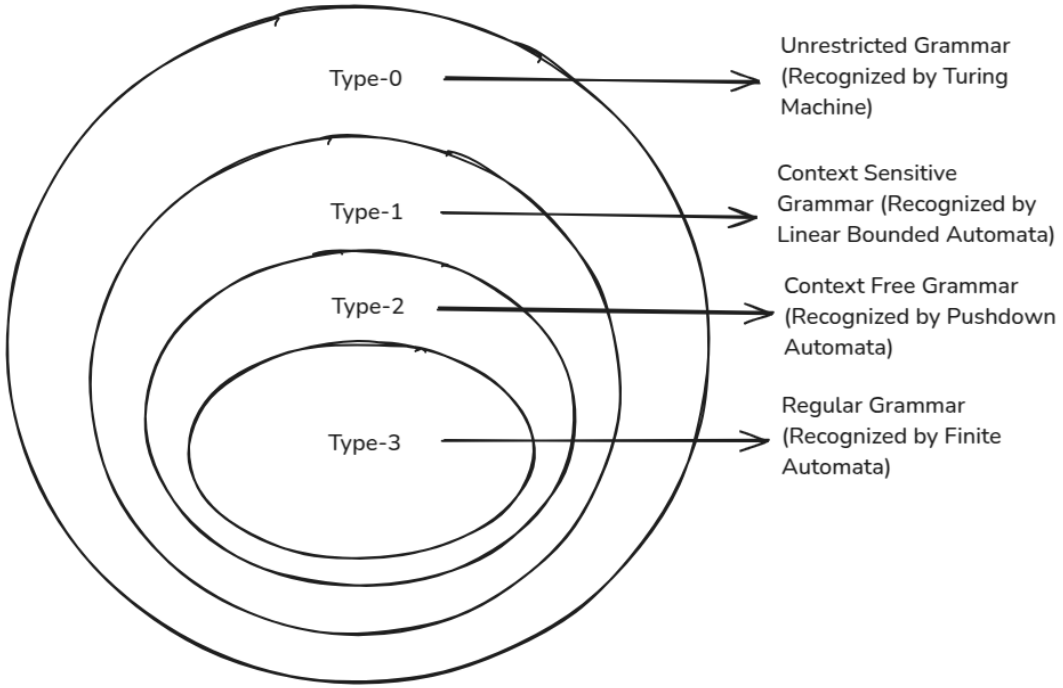
Context:
[1] tag contains tag.
[2] tag contains 5 pairs of tags.
[3] The text within 1st tag is “Home”
. . .
[7] The text within 5th tag is “Login/Logout”

Question: How to login into my user account?

Disambiguations:
DQ 1: What is the operation required?
DA 1: Login
DQ 2: Where is the login button in the given webpage?
DA 2: In the navbar
DQ 3: Where is the login button within the navbar?
DA 3: 5th tag within the navbar.

Answer: The user can login into their account by scrolling up on the navbar and clicking the fifth option in the list, namely “Login/Logout”.

Results – Qualitative (Contd.)



Chomsky Hierarchy

- From Chomsky Hierarchy, we observe that the most abstract class of languages is the one identified by a Turing machine.
- From Church Turing Thesis, all computable functions that exist can be read by a Turing machine.
- Effectively, any logic in principle for which we can define a computable function can be accepted by our language specification system.

Conclusion & Future Works

Conclusions

- We were able to generalize the concept of a machine learning model into a mathematical framework.
- We were able to apply this mathematical framework in the form of an abstract data structure.
- Such a framework can in principle work upon the class of all computable functions.
- We can use this framework as a design guideline to build an AI enabled software system that is extendible and easier to maintain.
- It can also be used to formally verify and analyze a machine learning project during its complete lifecycle.

Future Works

- Explore alternatives to point cloud representation.
- Explore distributed approach to the framework.
- Explore domain specific performance.
- Evaluate explainability.
- Perform and gather more insights on more rigid performance benchmarks.
- Explore possible optimizations.

Thank You